Multi-resolution Mapping Using Surface, Descent and Orbital Images

Clark F. Olson, Larry H. Matthies, Yalin Xiong
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099

Rongxing Li, Fei Ma, Fengliang Xu
Department of Civil and Environmental Engineering and Geodetic Science
The Ohio State University, Columbus, OH 43210-1275

Abstract

Our objective is to produce high-accuracy maps of the terrain elevation at landing sites on planetary bodies through the use of all available image data. We use images on the planetary surface from landers and rovers, images captured during the lander descent to the surface, and orbital images. Three new capabilities have been developed. First, we generate elevation maps from descent images using structure-from-motion techniques. These maps are useful for rover navigation and provide a link between the orbital images and surface images. We have developed a methodology for performing rover localization using bundle adjustment that uses tie points between the rover and descent images to determine both the camera and the tie point locations. Finally, a new method to perform registration between orbital images and descent images has been developed that locates the landing position in the orbital imagery and allows integration of the entire data set. These technologies are important for performing rover navigation in future space missions and the maps provide a tool for coordinating rovers in a robotic colony.

Keywords: terrain mapping, structure-from-motion, rover localization, image registration

1 Introduction

For the exploration of planetary surfaces, the images taken during the lander's descent provide a critical link between orbital imagery and imagery taken on the surface using rovers and landers. The descent imagery not only provides information about the landing location in a global coordinate system, but also yields progressively higher resolution maps for mapping and mission planning for rovers. We address the issue of mapping using all available imagery, including imagery from the surface, descent imagery, and orbital imagery.

In order to map the data from all sources of imagery and combine them into a multi-resolution map, we have developed several new capabilities. First, we have developed techniques for building three-dimensional terrain maps from descent imagery by comparing each pair of images in the nested sequence. Next, we have created techniques for localizing rovers on the surface using the descent imagery. This allows the incorporation of surface imagery into the multi-resolution map structure. Finally, we developed a new method for the registration of descent imagery to orbital imagery using entropy alignment.

Our approach to mapping descent imagery has two steps: motion refinement and depth recovery. In motion refinement, we use an initial motion estimate to avoid the intrinsic ambiguity in descending motions. The objective of the motion refinement is to adjust the motion parameters such that the epipolar constraints are valid between adjacent frames. The depth recovery step correlates adjacent frames to match pixels for triangulation. Due to the descending motion, the conventional rectification process is replaced by a set of anti-aliasing image warpings corresponding to a set of virtual parallel planes.

In order to locate rover positions in the map constructed from descent imagery and enable incorporation of the rover imagery into the map, we have developed bundle adjustment techniques for rover localization. In this method tie points that represent the same locations in the rover and descent imagery and determined and optimization techniques are used to determine the camera and tie point positions for each image in a batch optimization. These techniques have been extended to allow incremental localization, so that new imagery can be efficiently added to the

network of tie points.

We have also developed a new method for comparing descent imagery to orbital imagery in order to locate the landing site and provide context for the descent imagery. Unlike matching using mutual information, which fails in our test cases, our method takes advantage of shape information in the imagery by comparing the alignment of spatial entropy in the orbital image and the descent image. The basic method is to transform each image into a new image representing the entropy at each location. The images are then compared (for example, using normalized correlation) to determine a position where the entropies are best aligned.

Each of these methods has been tested by applying it to a set of data collected during rover field testing at Silver Lake, California. During this field test, a set of descent imagery was collected using a helicopter. The data set consists of eight images taken at elevations that range from 1086 m to 8 m above the ground. These images, together with the thousands of rover images collected at the site and SPOT satellite imagery, yield a rich set of data for testing rover mapping and localization algorithms in the context of a planetary landing scenario.

2 Mapping Descent Images

We recover depth maps from descent images using a two-stage process [11]. First, motion refinement is performed in order to guarantee that the epipolar constraints are satisfied between the images. Then, depth recovery is performed using an algorithm based on correlation.

2.1 Motion Refinement

Recovering camera motion from two or more frames is one of the classical problems in computer vision. Linear [5] and nonlinear [8] solutions have been proposed. For descent motions (as in Fig. 1), generic motion recovery from matched features is ill-posed owing to a numerical singularity. Since the camera is rigidly attached to the lander, and the change in the lander orientation can be measured accurately by an inertial navigation system onboard, we can eliminate the singularity problem by adding a penalty term for deviating from the measured orientation. We recover the camera motion by tracking features in the image and optimizing based on epipolar constraints.

For each pair of adjacent frames in the sequence, we track features that have been selected in the higher

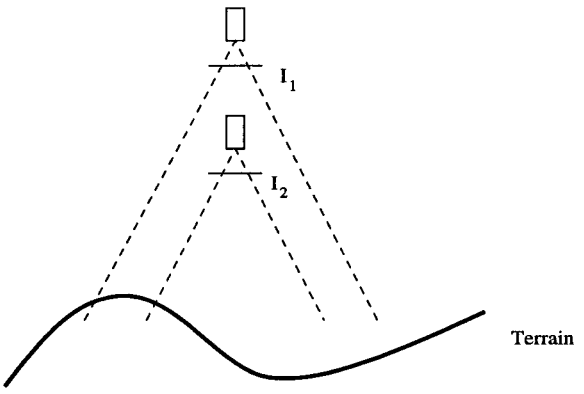


Figure 1: Descent Motion

resolution frame into the lower resolution frame. We use Forstner's interest operator [2] to evaluate the trackability of the features in the higher resolution frame. We select the features with high scores, while disallowing features that are too close together. Once the image resolutions have been equalized (through downsampling, or anti-aliasing warping, if necessary), feature tracking is performed in a straightforward manner using normalized correlation.

The tracked features provide a rich set of observations to constrain the camera motion, even though the relationship between the locations of the tracked features and the camera motion parameters is highly nonlinear. Let us assume that the projection matrix of the camera is $\mathbf{M}$, the location of feature i at time t is $[X_i^t, Y_i^t, Z_i^t]^T$, its image location at time t represented in homogeneous coordinates is $[x_i^t, y_i^t, z_i^t]^T$, and the camera motion between time t and time $t+1$ is composed of a translation $\mathbf{T}$ and rotation $\mathbf{R}$ (3×3 matrix). The projection of the feature at time t is, thus:

$$\begin{bmatrix} x_i^t \\ y_i^t \\ z_i^t \end{bmatrix} = \mathbf{M} \begin{bmatrix} X_i^t \\ Y_i^t \\ Z_i^t \end{bmatrix}, \tag{1}$$

and the projection at time $(t+1)$ is:

$$\begin{bmatrix} x_i^{t+1} \\ y_i^{t+1} \\ z_i^{t+1} \end{bmatrix} = \mathbf{M} \begin{bmatrix} X_i^{t+1} \\ Y_i^{t+1} \\ Z_i^{t+1} \end{bmatrix} = \mathbf{M} \left(\mathbf{R} \begin{bmatrix} X_i^t \\ Y_i^t \\ Z_i^t \end{bmatrix} + \mathbf{T} \right). \tag{2}$$

Therefore, the feature motion in the image is:

$$\begin{aligned} \begin{bmatrix} x_i^{t+1} \\ y_i^{t+1} \\ z_i^{t+1} \end{bmatrix} &= \mathbf{M} \left(\mathbf{R}\mathbf{M}^{-1} \begin{bmatrix} x_i^t \\ y_i^t \\ z_i^t \end{bmatrix} + \mathbf{T} \right) \\ &= \mathbf{U} \begin{bmatrix} x_i^t \\ y_i^t \\ z_i^t \end{bmatrix} + \mathbf{V}, \end{aligned} \tag{3}$$

where $\mathbf{U} = \mathbf{MRM^{-1}}$ is a 3×3 matrix and $\mathbf{V} = \mathbf{MT}$ is a 3-vector. Let $[c_i^t, r_i^t] = [x_i^t/z_i^t, y_i^t/z_i^t]$ denote the actual column and row location of feature i in image coordinates at time t. We, then, have the predicted feature locations at time $t+1$ as:

$$\hat{c}_i^{t+1} = \frac{u_{00}x_i^t + u_{01}y_i^t + u_{02}z_i^t + v_0}{u_{20}x_i^t + u_{21}y_i^t + u_{22}z_i^t + v_2}, \quad (4)$$

$$\hat{r}_i^{t+1} = \frac{u_{10}x_i^t + u_{11}y_i^t + u_{12}z_i^t + v_1}{u_{20}x_i^t + u_{21}y_i^t + u_{22}z_i^t + v_2}, \quad (5)$$

where u_{ij} and v_i are elements of $\mathbf{U}$ and $\mathbf{V}$ respectively.

In order to refine the motion estimate, we augment the parameters with depth estimates for each of the features. There are two advantages to this approach. First, the objective function becomes strictly the distance between the predicted and observed feature locations. Therefore, it is guaranteed to have no bias if the observations contain no bias. In addition, in the context of mapping descent images, we have a good initial estimate of the depth value from the spacecraft altimeter. Incorporating this information will, thus, improve the optimization in general.

Let us say that the depth value of feature i at time t is d_i^t and the camera is pointing along the z-axis, the homogeneous coordinates of the feature are $[x_i^t, y_i^t, z_i^t]^T = d_i^t[c_i^t, r_i^t, 1]^t$. Therefore, the overall objective function we are minimizing is:

$$\sum_{i=0}^{N-1} \left(\left(c_i^{t+1} - \hat{c}_i^{t+1}\right)^2 + \left(r_i^{t+1} - \hat{r}_i^{t+1}\right)^2 \right), \quad (6)$$

where N is the number of features, and $\hat{c}_i^{t+1}$ and $\hat{r}_i^{t+1}$ are nonlinear functions of the camera motion and depth value d_i^t given by Eq. (4) and (5). We perform nonlinear minimization using the Levenberg-Marquardt algorithm.

2.2 Depth Map Recovery

The second step of our method generates depth maps using correlations between image pairs. In order to compute the image correlation efficiently, we need to rectify the images in a manner similar to binocular stereo. Unfortunately, it is impossible to rectify the images along scanlines because the epipolar lines intersect each other near the center of the images. If we resample the images along epipolar lines as in stereo, we will oversample near the image center, and undersample near the image boundaries.

In order to avoid this problem, we adopt a slicing algorithm that allows us to perform the correlation efficiently. The main concept is to use a set of virtual planar surfaces slicing through the terrain. The virtual planar surfaces are similar, in concept, to horopter surfaces [1] in stereo. For every planar surface k, if the terrain surface lies on the planar surface, there exists a projective warping $\mathbf{P}_k$ between two images. If we designate the first image $I_1(x, y)$ and the second image $I_2(x, y)$, then for every virtual planar surface, we can compute a correlation image as the sum-of-squared-differences (SSD):

$$C_k(x,y) = \sum_{m=x-W}^{x+W} \sum_{n=y-W}^{y+W} \left(I_1(m,n) - I_2^k(m,n)\right)^2, \quad (7)$$

where $2W+1$ is the size of the correlation window and $I_2^k(x, y)$ is a warped version of $I_2(x, y)$:

$$I_2^k(x,y) = I_2\left(\frac{p_{00}x + p_{01}y + p_{02}}{p_{20}x + p_{21}y + p_{22}}, \frac{p_{10}x + p_{11}y + p_{12}}{p_{20}x + p_{21}y + p_{22}}\right), \quad (8)$$

and p_{ij} are elements of the 3x3 matrix $\mathbf{P}_k$. Due to the drastic resolution difference, an anti-aliasing resampling such as [3] or a uniform downsampling of $I_2(x, y)$ is applied before the image warping. In practice, if the camera heading directions are close to be perpendicular to the ground, a uniform downsampling before warping shall suffice. Otherwise, a space-variant downsampling is needed to equalize the image resolutions.

The depth value at each pixel is the depth of the planar surface z_k whose corresponding SSD image pixel $C_k(x, y)$ is the smallest:

$$z(x, y) = z_k, \quad (9)$$

where

$$C_k(x, y) \leq C_j(x, y), j = 0, \ldots, M-1, \quad (10)$$

and M is the number of planar surfaces. To further refine the depth values, the underlying SSD curve can be interpolated by a quadratic curve and the "subpixel" depth value can be computed [10] as:

$$z(x,y) = z_k + \frac{\delta z(C_{k+1}(x,y) - C_{k-1}(x,y))}{2(C_{k+1}(x,y) + C_{k-1}(x,y) - 2C_k(x,y))}, \quad (11)$$

where δz is the depth increment between adjacent planar surfaces.

The projective warping matrix $\mathbf{P}_k$ is derived from the parameters of the camera motion and the planar surfaces. For an arbitrary point $\mathbf{X}$ in some reference frame, its projection is expressed as $\mathbf{x} = \mathbf{M}(\mathbf{X} - \mathbf{C})$, where $\mathbf{C}$ is the position of the camera nodal point and $\mathbf{M}$ is the projection matrix. Note that $\mathbf{C}$ and

$\mathbf{M}$ encapsulate the camera motion between the images, since they are represented in a common reference frame. Let $\mathbf{C}_1$ and $\mathbf{M}_1$ represent the higher camera, $\mathbf{C}_2$ and $\mathbf{M}_2$ represent the lower camera, and $\mathbf{N}^T\mathbf{X} + z_k = 0$ represent the set of planar surfaces. For any pixel in image 2 (i.e. the lower camera), its location must lie on a 3d ray:

$$\mathbf{X} = s\mathbf{M}_2^{-1} \begin{bmatrix} c_2 \\ r_2 \\ 1 \end{bmatrix} + \mathbf{C}_2, \tag{12}$$

where c_2 and r_2 are the column and row location of the pixel and s is a positive scale factor. If the pixel is from a point on the planar surface, then the following constraint must be satisfied:

$$s\mathbf{N}^T\mathbf{M}_2^{-1} \begin{bmatrix} c_2 \\ r_2 \\ 1 \end{bmatrix} + \mathbf{N}^T\mathbf{C}_2 + z_k = 0. \tag{13}$$

Therefore, the scale factor s must be

$$s = -\frac{\mathbf{N}^T\mathbf{C}_2 + z_k}{\mathbf{N}^T\mathbf{M}_2^{-1}[c_2, r_2, 1]^T}. \tag{14}$$

We can then re-project the point onto the first image using Eq. (12) and (14):

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \mathbf{M}_1(\mathbf{X} - \mathbf{C}_1) = \mathbf{P}_k \begin{bmatrix} c_2 \\ r_2 \\ 1 \end{bmatrix}, \tag{15}$$

where $\mathbf{P}_k$ is a 3x3 matrix specifying the projective warping:

$$\mathbf{P}_k = \mathbf{M}_1(\mathbf{C}_2 - \mathbf{C}_1)\mathbf{N}^T\mathbf{M}_2^{-1} - (\mathbf{N}^T\mathbf{C}_2 + z_k)\mathbf{M}_1\mathbf{M}_2^{-1}. \tag{16}$$

Note that the depth recovery is numerically unstable in the vicinity of the epipoles (at the center of the image for pure descending motion), since pixels near the epipoles have a small amount of parallax, even for large camera motions. Mathematically, the SSD curves in those areas are very flat and, thus, accurate depth recovery is difficult. These regions can be easily filtered, if desired, by imposing a minimum curvature threshold at the minima of the SSD curves.

2.3 Experiments

A set of real descent images was collected in the desert near Silver Lake, California using a helicopter. Figure 2 shows four frames from this sequence. The initial camera motions were estimated using control points on the ground. Several of the images contain significant lateral motions due to the difficulty in maintaining the x-y position of the helicopter during the data collection. Column (c) of Fig. 2 shows the false-color depth maps that were recovered from the sequence and column (d) shows the image draped over the visualized terrain.

For the images in this data set, the terrain slopes downward from left to right, which can be observed in the rendered maps. Some of the interesting terrain features include the bushes visible in row 1 and the channel in row 2. Note that the areas in which the helicopter shadow is present yield good results, despite the movement of the shadow. This can be attributed to the robust methods that we use for both motion estimation and template matching [7]. Overall, this data set indicates that we can robustly compute maps that are useful for rover navigation over both small and large scales using real descent images.

3 Localization With Descent Images

A second aspect of the multi-resolution mapping problem that we have examined is the determination of the rover/camera position in the terrain using matches between surface and descent images [4]. This allows the construction of a map encompassing both sets of images.

For landing on Mars (or another planetary body), we have no ground control points. In this case, the bundle adjustment computation is a free network consisting of the exposure centers of the camera positions (both descent and surface images), measured image tie points, and the ground location of each of the tie points. We select the landing location as the origin of a local coordinate frame. Three constraints are applied to the bundle adjustment model: scale, azimuth, and zenith. These are supplied by a landmark location relative to the lander obtained by, for example, stereo vision. Given the landmark coordinates (L_x, L_y, L_z), we constrain the scale S, azimuth α, and zenith β as follows:

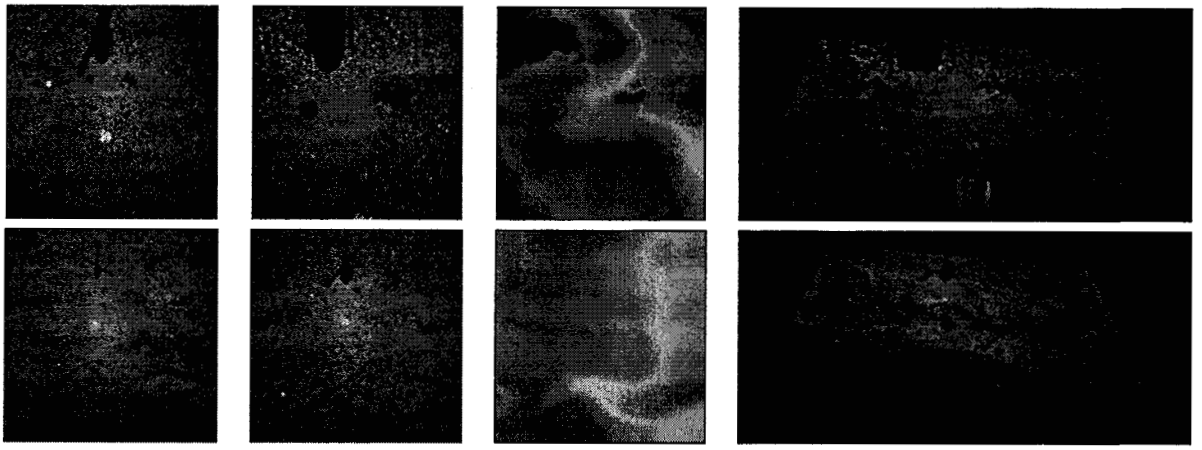
$$S^2 = {L_x}^2 + {L_y}^2 + {L_z}^2 \tag{17}$$

$$\alpha = \tan^{-1} L_y/L_x \tag{18}$$

$$\beta = \tan^{-1} L_z/\sqrt{{L_x}^2 + {L_y}^2} \tag{19}$$

If we let A be the coefficient matrix after linearization, L be the observation vector and V be the correction vector, then we have the unknown vector X (including the camera positions and ground coordinates) as:

$$V = AX - L, \tag{20}$$



(a) (b) (c) (d)

Figure 2: Real descent sequence from a helicopter. (a) Image taken at higher altitude. (b) Image taken at lower altitude. (c) False-color estimated terrain map. (d) Rendered terrain map with image overlaid. (The rows have different height scales.)

and the three constraints above can be represented as:

$$HX = W. \tag{21}$$

With weight matrix P on the measurements of the image points, we let $N = A^T PA$ and our least-squares solution becomes:

$$X = N^-(A^T PL + H^T(HN^-H^T)^{-1}(W - HN^-A^T PL)), \tag{22}$$

where N^- is the generalized inverse of N.

For rover localization along a traverse, we would prefer an incremental method, rather than a batch method that may require more time than desired. Assuming that we have previously processed the first $m-1$ rover stations, we can decompose the observation equations into two parts:

$$v_{m-1} = A_{m-1}X_{m-1} - l_{m-1}, \tag{23}$$

and

$$v_m = A_m X_m + B_m Y_m - l_m. \tag{24}$$

Equation 23 represents the solution with all of the data until rover station $m-1$ and Equation 24 is an incremental update of the position using the data at station m, where Y_m is the new unknown vector expressing the rover position at station m. In this case, the generalized inverse of N_m becomes:

$$N_m^- = \begin{bmatrix} A_{m-1}^T P_{m-1} A_{m-1} + A_m^T P_m A_m & A_m^T P_m B_m \\ B_m^T P_m A_m & B_m^T P_m B_m \end{bmatrix}^- \tag{25}$$

$$= \begin{bmatrix} K_m & G_m \\ G_m^T & H_m \end{bmatrix}. \tag{26}$$

The incremental solution for station m is:

$$X_m = \hat{W}_m(X_{m-1} - F_m(l_m - A_m X_{m-1})) + G_m B_m^T P_m l_m \tag{27}$$

and

$$Y_m = -(B_m^T P_m B_m)^- B_m^T P_m A_m \hat{W}_m \cdot (X_{m-1} - F_m(l_m - A_m X_{m-1})) + H_m B_m^T P_m I_m \tag{28}$$

where

$$\hat{W}_m = I + \hat{N}_m^- A_m^T P_m B_m H_m B_m^T P_m A_m^T, \tag{29}$$

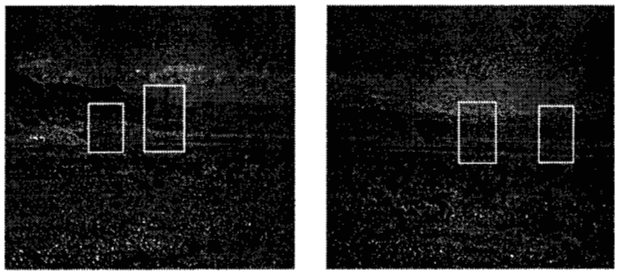
$$F_m = (A_{m-1}^T P_{m-1} A_{m-1})^- A_m^T \cdot (P_m^{-1} + A_m(A_{m-1}^T P_{m-1} A_{m-1})^- A_m^T)^{-1}, \tag{30}$$

$$\hat{N}_m = (N_{m-1} + A_m^T P_m A_m)^-. \tag{31}$$

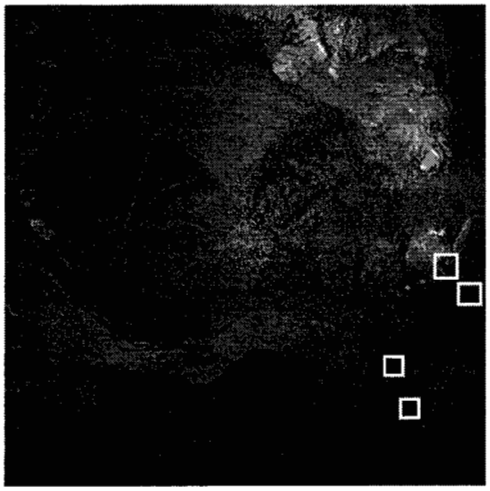
This method was tested using data from the Silver Lake test site. In this experiment, we used 11 descent images and 14 pairs of rover stereo images taken at 3 separate rover stations. Image features appearing in both the rover and descent images were selected as tie points (see Fig. 3). The results of the experiment were compared with ground-truth collected using GPS. In each case, we were able to obtain localization accuracy within 1 m of the GPS estimate. For rover positions closer to the center of the descent imagery, the accuracy is much better, with errors below 5 cm for a position approximately 5 meters from the center.

4 Registering Descent Images

In order to determine the location of the landing site and provide context for the descent images, it is



(a)



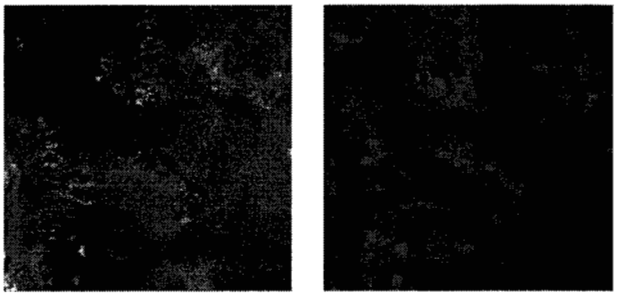
(b)

Figure 3: Example tie points between rover and descent images. (a) Rover images. (b) Descent image.

necessary to determine the location of the descent images within lower-resolution orbital images taken of the same location. This is a difficult problem for two reasons. First, the images are captured with different sensors that have different sensitivities to various wavelengths of light. Therefore, the same terrain location will yield different image intensities for the two sensors. In general, the relationship between the image intensities yielded by the same terrain location in the two images is highly non-linear. The second problem is that transformation between the camera positions (and, thus, the position of the image data) is complex. There are six degrees-of-freedom in the relative camera positions and this leads to a six degree-of-freedom transformation in the image space, if the terrain is approximated as planar.[1]

A common technique that is used for this type of image registration is the maximization of mutual information between the images [6, 9]. This technique locates the relative position between the images at which the mutual statistical information content is maximized. Unfortunately, in experiments on real images, we have found that this method fails when the search space is large. We speculate that the reason for this failure is that mutual information does not well use shape information that is present in the images. Another possible explanation is that smooth shading from the different illumination in the images causes the correct match to score poorly, since mutual information can not handle this type of illumination change.

[1]The transform is even more complex if the terrain is not approximated as planar. However, we shall use the planar approximation in this paper, since the distance of the terrain from the camera is large.



(a) (b)

Figure 4: Entropy image example. (a) Orbital image of the Ayawatz Mountains and Silurian Valley in California. (b) Entropy image computed from (a).

We use a different method, where each image is transformed into an entropy image, storing the entropy at each location in the original image. Registration is then performed using the entropy images. This method is more robust to changes in illumination and makes greater use of shape information in the image.

For a discrete random variable A, with marginal probability distribution $p_A(a)$, the entropy is defined as:

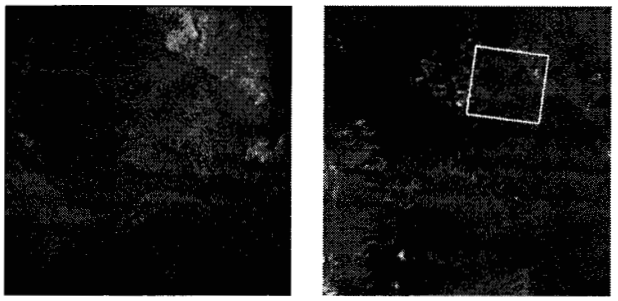
$$H(A) = -\sum_a p_A(a) \log p_A(a). \tag{32}$$

Note that $0 \cdot \log 0$ is taken to be zero, since

$$\lim_{x \to 0} x \log x = 0. \tag{33}$$

In order to compute the entropy image for both the template (the descent image) and the search image (the orbital image), we apply Eq. 32 to each square window of some particular size (15×15 pixels is typical) and replace the pixel value with the entropy score. Figure 4 shows an example of an entropy image created from an orbital image at our test site.

In order to locate the best position of the descent image in the orbital image, we combine the use of the fast Fourier transform (FFT) to perform correlation



(a) (b)

Figure 5: Registration example. (a) Aerial image showing a detail of the image in Fig 4. (b) Registered location of the aerial image with respect to the orbital image.

over translations with a search over the remaining four parameters of the search space. For each set of rotation, scaling, and warping parameters, correlation is performed efficiently in the frequency domain and the location with the highest normalized correlation can be located very quickly. At present, a brute-force search is used to search over the remaining parameters of the search space. We are currently investigating algorithms for efficiently searching these pose parameters. Fig. 5 shows an example of the registration achieved with these techniques. In this example (where experiments with mutual information have failed to detect the correct position), the match using entropy alignment performs well, finding a very close match between the descent image and the orbital image.

5 Summary

We have discussed new techniques for performing multi-resolution mapping of image data from a variety of sources. For mapping landing sites on planetary bodies, we use images taken on the surface by landers and rovers, images taken during a lander's descent to the surface, and images captured from orbit. Three techniques were described in this paper. First, a method for generating terrain maps from the descent images was described. This gives us multi-resolution information for navigation and planning, and provides a link between the surface images and the orbital images. Next, we described a method for determining the rover position on the planet using correspondences between surface imagery and descent imagery. In addition to being useful for navigation, these techniques allow the surface imagery to be accumulated into a map encompassing all of the data. Finally, we have described techniques for performing registration between descent images and orbital images. This allows us to determine the location of the landing site and provides context for the descent images and, by extension, the surface images. The combination of these techniques yields an overall framework for registering and combining the terrain maps for all of the data sources.

Acknowledgments

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] P. J. Burt, L. Wixson, and G. Salgian. Electronically directed "focal" stereo. In *Proceedings of the International Conference on Computer Vision*, pages 94–101, 1995.

[2] W. Förstner. A framework for low-level feature extraction. In *Proceedings of the European Conference on Computer Vision*, pages 383–394, 1994.

[3] P. Heckbert. Survey of texture mapping. *IEEE Computer Graphics and Applications*, 6(11):56–67, November 1986.

[4] R. Li, F. Ma, F. Xu, L. Matthies, C. Olson, and Y. Xiong. Large scale Mars mapping and rover localization using descent and rover imagery. In *Proceedings of the 19th ISPRS Congress*, 2000.

[5] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[6] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.

[7] C. F. Olson. Maximum-likelihood template matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 52–57, 2000.

[8] R. Szeliski and S. B. Kang. Recovering 3d shape and motion from image streams using non-linear least squares. *Journal of Visual Communication and Image Representation*, 5(1):10–28, March 1994.

[9] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[10] Y. Xiong and L. H. Matthies. Error analysis for a real-time stereo system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1087–1093, 1997.

[11] Y. Xiong, C. F. Olson, and L. H. Matthies. Computing depth maps from descent imagery, 2001. Submitted to the *IEEE International Conference on Computer Vision*.